ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul246

Academic Risk Profiling and Tailored Student Guidance Through Machine Learning in Kinshasa Universities

Augustin Pambi Tadiamba[1*]; Pierre Katalay Kafunda[2]; David Mayoya Kutangila[3]; Joseph Shanga Minga[4]; Bebe Bebe Ngwaba[5]; Joseph Bimbala Ngwaba[6]; Grace-Roven Gracia Tshimanga[7]; Jean-Jacques Matondo Katshitshi[8]; Roger Kwenge Mboma[9]

Teaching Assistant and Research[1,5,9]; Full Professor and Researcher[2]; Professor and Rsearcher[3]; Senior Lecturer and Researcher[4,6,7,8]

[1,3,4,5,6,7,8,9]Department of Computing, Management and English for Businesses, University of Kinshasa, Kinshasa, Democratic Republic of the Congo
[2]Department of Mathematics, Cmpouter Science and Statistics, Faclty of Science an Technoogy, Univrsity of Kinshasa,Kinshasa, Democratic Republic of the Congo

[1]Corresponding Author: Augustin Pambi Tadiamba[1*]

Publication Date: 2026/07/22

Abstract: Educational data can help identify, before the end of a semester, academic trajectories that deserve timely attention. This study develops an academic risk profiling and tailored student guidance using machine learning techniques for universities in Kinshasa. A quantitative, experimental, and predictive design was applied to a harmonized dataset of 9,000 student-semester observations and fourteen predictors after removing a redundant composite engagement index. Data preparation, modelling, and visualisation were performed in a reproducible Python/Jupyter Notebook workflow using pandas, NumPy, scikit-learn, and Matplotlib. A stratified 80/20 training-test split and five-fold stratified cross-validation were used to compare multinomial logistic regression, decision tree, random forest, and multilayer perceptron models. On the independent test set, the multilayer perceptron achieved the highest accuracy (0.720) and macro-AUC-ROC (0.945), while logistic regression achieved the highest balanced accuracy (0.714). Random forest was retained because it achieved the highest macro F1 score (0.686), the prespecified criterion for protecting attention to minority classes, together with macro precision of 0.713 and macro One-vs-Rest AUC-ROC of 0.941. Continuous assessment average, midterm average, and the previous validated-credit rate were the most informative signals. Recommendations are derived from the predicted class, while uncertain cases are flagged for human review. The proposed model is therefore a decision-support tool rather than an automated decision-maker.

Keywords: *Machine Learning; Educational Data; Academic Risk Profiling; Random Forest; Tailored Student Guidance.*

How to Cite: Augustin Pambi Tadiamba; Pierre Katalay Kafunda; David Mayoya Kutangila; Joseph Shanga Minga; Bebe Bebe Ngwaba; Joseph Bimbala Ngwaba; Grace-Roven Gracia Tshimanga; Jean-Jacques Matondo Katshitshi; Roger Kwenge Mboma (2026) Academic Risk Profiling and Tailored Student Guidance Through Machine Learning in Kinshasa Universities. *International Journal of Innovative Science and Research Technology*, 11(7), 617-624. https://doi.org/10.38124/ijisrt/26jul246

I. INTRODUCTION

Higher education institutions generate records that describe student progress: registrations, assessment results, assignments, quizzes, attendance, credits earned, and, increasingly, traces from digital learning environments. When these records are analysed responsibly, they can support pedagogical decisions instead of serving only administrative purposes [1, 2, 3, 4]. Academic-performance prediction is one of the most widely studied applications of educational data mining and learning analytics. Supervised models can combine academic and behavioural attributes to estimate a later outcome, while early-alert systems make it possible to identify students who may need support before difficulties

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul246

become persistent [5, 6, 7, 8]. In a university setting, however, a prediction has limited value if it is reduced to a score or label. It must be interpreted in context and translated into a support option that remains open to professional judgement.

Universities in Kinshasa already maintain a substantial portion of the information required for this type of analysis. In practice, such data are often fragmented, examined late, and rarely organised as semester-level alerts. The challenge is therefore twofold: to assess six performance levels fairly despite class imbalance and to provide a pedagogical response without allowing the algorithm to replace academic judgement. This study addresses that challenge by evaluating early prediction models and linking their outputs to a graded academic recommendation framework.

➢ *Contributions to Practice and Theory*

The study contributes a six-level early-warning formulation—Risk, Fragile, Average, Good, Very Good, and Excellence—rather than a simplified pass/fail classification. It also uses predictors that are observable by the middle of the semester, evaluates competing models with a metric suited to minority classes, and converts the selected model output into a transparent recommendation rule. For practice, the work offers a structured basis for identifying students who may need tutoring, feedback, consolidation, enrichment, or a closer human review. For theory, it connects predictive performance with interpretability, uncertainty signalling, and the educational use of learning-analytics outputs [9, 10, 11, 12, 13].

II. LITERATURE REVIEW

➢ *Educational Data and Early Prediction*

Learning analytics and educational data mining have shifted attention from describing past academic results to identifying patterns that can inform timely action [1, 2, 3, 4]. Prior studies show that assessments, prior achievement, attendance, and indicators of participation can contribute to performance prediction [5], [6]. Early-warning initiatives consequently focus on signals that become available while intervention remains possible, rather than after a semester has already ended [7], [8].

➢ *Model Selection, Evaluation, and Reproducibility*

No algorithm is universally superior for academic prediction. Linear models offer transparent baselines, decision trees expose decision rules, random forests capture nonlinear interactions, and multilayer perceptrons can learn more complex relationships [14], [15]. When the target classes are unevenly represented, a global accuracy score can obscure weak recognition of less common profiles. Macro F1, balanced accuracy, class-level measures, and One-vs-Rest AUC-ROC therefore provide a more informative evaluation set [16].

➢ *Interpretation, Recommendations, and Human Oversight*

Learning-analytics outputs should be used to guide dialogue and support, not to impose irreversible decisions. The ethical literature stresses transparency, confidentiality, data minimisation, and human oversight when student data are used for prediction [11, 12, 13]. This study therefore keeps the recommendation rule explicit: the predicted class proposes a type of support, while low-confidence or low-margin cases are directed to a teacher or academic authority for review.

III. MATERIALS AND METHODS

➢ *Research Design and Experimental Corpus*

This research follows a quantitative, experimental, and predictive design. The unit of analysis is the student-semester observation: each record describes a student situation observed after the mid-semester point, whereas the target represents the final performance level reached at the end of that same semester. The standardized and anonymized corpus supports methodological validation; it is not intended as an institutional prevalence estimate.

The dataset comprises 9,000 observations drawn from three universities represented in the experimental corpus: the University of Kinshasa (4,105; 45.6%), the National Pedagogical University (3,042; 33.8%), and the Protestant University of Congo (1,853; 20.6%). The target distinguishes Risk, Fragile, Average, Good, Very Good, and Excellence. The Risk (5.2%) and Excellence (2.1%) classes are less frequent; their recognition was therefore given particular attention during evaluation.

➢ *Variables and Data Preparation*

Fifteen conceptual variables were initially selected from student records and pedagogically observable indicators. The academic engagement index was removed because it aggregates attendance, participation, and assignment-submission information already represented separately in the matrix. The final matrix contains fourteen predictors: eleven numerical and three categorical. This choice reduces redundancy and makes the contribution of each educational indicator easier to interpret.

Table 1 Groups of Variables Retained for Early Modelling

Band	Main variables	Analytical role
Institutional context	University; programme/faculty; level of study	Situates the trajectory within the curriculum and controls the context.
Academic history	Academic history available; cumulative average; previously validated credits; failed courses	Describes the student's previous progress.
Mid-semester assessments	Midterm average; continuous assessments; assignments; quizzes	Provides early signals for the current semester.
Observable behaviours	Attendance; participation; assignment-submission rate	Helps interpret interim results with greater nuance.

Source: Authors' Development from the Harmonized Experimental Dataset (n = 9,000).

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul246

Missing numerical values were imputed with the median estimated from the training data only. Categorical values were imputed with the most frequent category and then one-hot encoded. Standardisation was applied to logistic regression and the multilayer perceptron, but not to tree-based models. These operations were placed within pipelines so that the test subset could not influence preprocessing or model estimation [17], [18].

➢ *Modelling and Experimental Protocol*

The task was formulated as a supervised multiclass classification problem. For each student-semester observation i, the model learns a function that associates the observed feature vector x_i with one of the six final performance classes $\hat{y}_i$. The target remains outside the predictors and is only observed at the end of the corresponding semester.

$$\hat{y}_i = f\theta(x_i) \quad (1)$$

The corpus was split into a stratified training subset of 7,200 observations (80%) and an independent test subset of 1,800 observations (20%), using random_state = 42. Five-fold stratified cross-validation was applied only to the training subset. The selected algorithms were chosen for their complementary qualities: a linear benchmark, a rule-based tree, an ensemble method able to model nonlinear interactions, and a neural network able to learn more complex patterns [14], [15].

Table 2 Algorithms and Main Experimental Parameters

Model	Main parameters	Role in comparison
Multinomial logistic regression	C = 1.0; L2 regularisation; balanced weighting; max_iter = 700	Interpretable and stable reference.
Decision tree	criterion = entropy; max_depth = 12; min_samples_leaf = 5; balanced weighting	Rule reading and threshold identification.
Random forest	220 trees; max_depth = 20; min_samples_leaf = 2; max_features = sqrt; balanced weighting	Robust treatment of nonlinear interactions.
Multilayer perceptron	Layers (64, 32); ReLU; alpha = 0.001; early stopping; max_iter = 350	Learning of more complex relationships.

Source: Authors' Development from the Harmonized Experimental Dataset (n = 9,000).

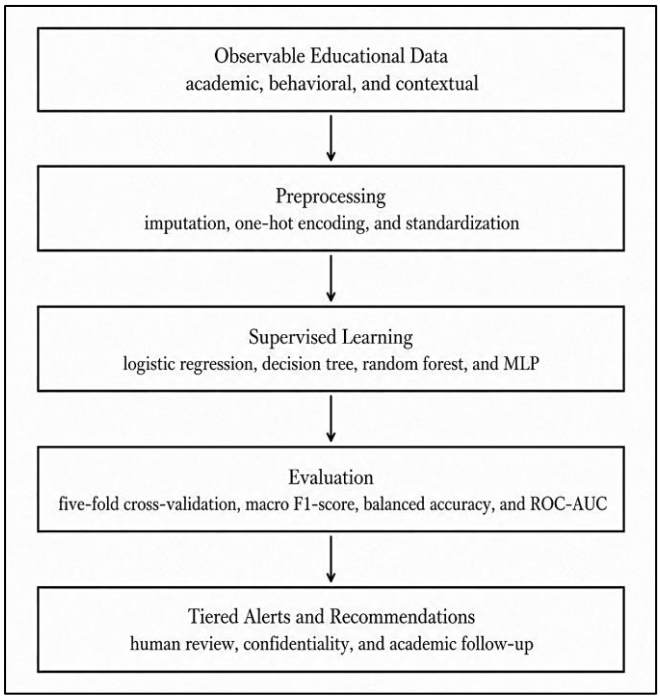

Fig 1 Experimental Early-Prediction and Academic-Recommendation Chain.
Source: Authors' Interpretation from Experimental Data.

➢ *Evaluation, Reproducibility, and Recommendation Rule*

Macro F1 was specified before training as the primary comparison criterion because it gives equal weight to every performance class, regardless of its frequency. This is particularly important in a six-class setting where Risk and Excellence are underrepresented. Macro F1 was complemented by overall accuracy, balanced accuracy, macro precision, class-level precision, recall, F1-score, a confusion matrix, and macro One-vs-Rest AUC-ROC when probability estimates were available [16].

$$F_{1,macro} = (1/K) \sum_{k=1}^{K} F_{1,k} \quad (2)$$

$$BA = (1/K) \sum_{k=1}^{K} Recall_k \quad (3)$$

Model selection was based on the mean macro F1 obtained through five-fold stratified cross-validation on the training subset. The independent test subset remained untouched until the final comparison; it was not used to tune parameters or select the retained classifier. The computational workflow was implemented in a Jupyter Notebook using Python. pandas supported data loading and organisation; NumPy supported numerical operations; scikit-learn provided preprocessing pipelines, model fitting, cross-validation, and evaluation; and Matplotlib produced the figures [19, 20, 21, 22, 23].

The recommendation rule was deliberately kept transparent. The predicted class determines the proposed intervention, $r_i = g(\hat{y}_i)$. Model probabilities do not alter the recommendation category. They are used only to show uncertainty: a case is flagged for educator review when the highest predicted-class probability is below 0.60 or when the margin between the first and second class probabilities is below 0.10. The teacher or academic authority may confirm, adjust, or reject the suggested intervention.

$$r_i = g(\hat{y}_i) \quad (4)$$

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul246

IV. RESULTS AND DISCUSSION

➢ *Cross-Validation and Independent Test Performance*

Five-fold cross-validation was conducted on the training subset only. Random forest obtained the highest mean macro F1 score (0.700; standard deviation = 0.023), followed by multinomial logistic regression (0.686; standard deviation = 0.015) and the multilayer perceptron (0.676; standard deviation = 0.032). Logistic regression recorded the highest mean balanced accuracy (0.742), whereas the decision tree remained lower on the principal indicators. Because macro F1 was set in advance as the model-selection criterion, random forest was retained for final evaluation.

Table 3 Five-Fold Stratified Cross-Validation Results on the Training Subset

Model	Macro F1 mean	F1 SD	Balanced accuracy mean	Accuracy mean
Random forest	0.700	0.023	0.688	0.724
Multinomial logistic regression	0.686	0.015	0.742	0.708
Multilayer perceptron	0.676	0.032	0.646	0.730
Decision tree	0.596	0.008	0.635	0.620

Source: Authors' Output from Five-Fold Stratified Cross-Validation on the Training Subset (n = 7,200).

The independent test results confirm that no algorithm dominates every indicator. The multilayer perceptron achieved the highest accuracy (0.720) and macro One-vs-Rest AUC-ROC (0.945), while logistic regression obtained the highest balanced accuracy (0.714). Random forest, however, produced the strongest macro F1 score (0.686) and macro precision (0.713). Its retention therefore follows the stated priority: recognising all six performance levels as evenly as possible rather than maximising only a global score that can be driven by frequent classes.

Table 4 Independent Test Performance of the Four Model Pipelines

Model	Accuracy	BA	Macro precision	Macro recall	Macro F1	Macro AUC
Random forest	0.711	0.665	0.713	0.665	0.686	0.941
Multilayer perceptron	0.720	0.650	0.697	0.650	0.670	0.945
Multinomial logistic regression	0.694	0.714	0.638	0.714	0.667	0.943
Decision tree	0.629	0.632	0.592	0.632	0.609	0.840

Source: Authors' Output from the Independent test Subset (n = 1,800). BA = Balanced Accuracy.

➢ *Influential Variables and Progressive Combination of Information*

Permutation importance measures the loss in macro F1 when the values of a predictor are randomly shuffled in the test data. The continuous-assessment average was the most influential variable (0.176), followed by the midterm average (0.162) and the previous validated-credit rate (0.137). Previous academic average, quizzes, assignments, and programme/faculty information also added predictive content. These results do not establish causal effects; they express each variable's relative predictive contribution within this dataset and the fitted random-forest model.

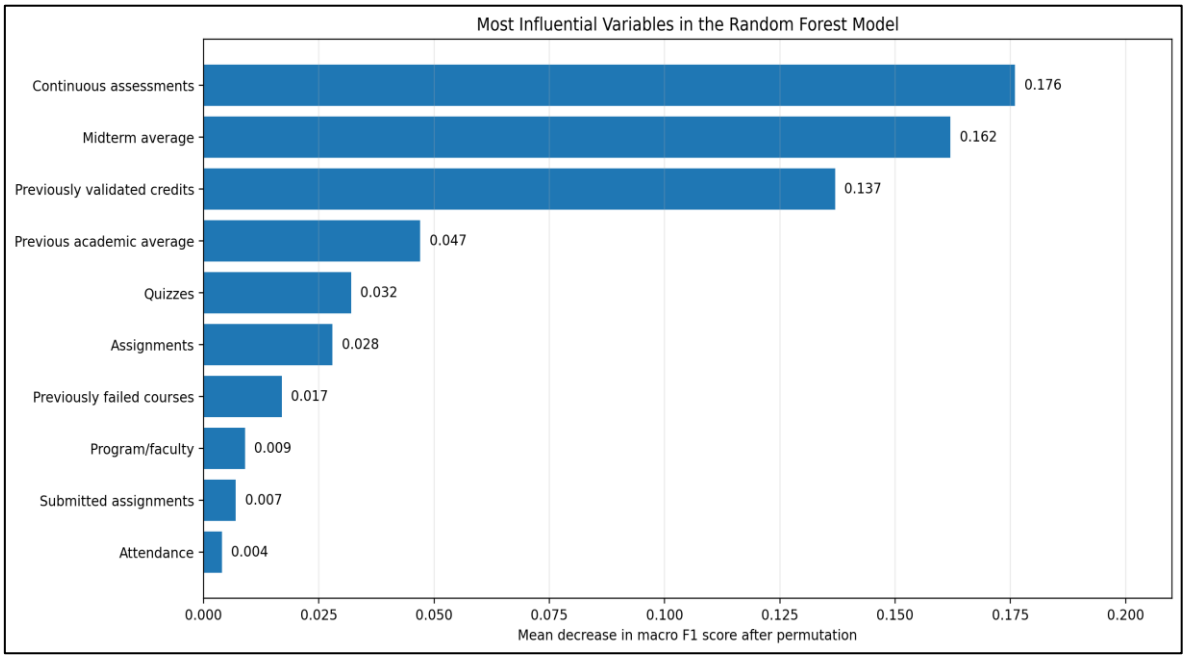

Fig 2 Permutation Importance of Variables in the Retained Random-Forest Model.
Source: Direct Graphical Output Generated by the Final Python/Jupyter Notebook.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul246

The scenario analysis evaluates the value of progressively enriching the student record. Using only four mid-semester academic indicators produced a macro F1 score of 0.621. Adding three behavioural indicators increased it to 0.629. The full set of academic, behavioural, contextual, and historical indicators reached 0.686, a gain of 0.065 over the first scenario. Early alerts should therefore not be based on one isolated grade; they become more credible when interim results are read together with prior trajectory and regularity indicators.

Table 5 Effect of Progressive Variable Combination on Test Performance

Scenario	No. variables	Accuracy	BA	Macro F1
A. Mid-semester academic indicators	4	0.649	0.605	0.621
B. Academic + behavioural indicators	7	0.653	0.606	0.629
C. Academic + behavioural + contextual/historical	14	0.711	0.665	0.686

Source: Authors' Scenario Analysis from the Final Random-Forest Workflow.

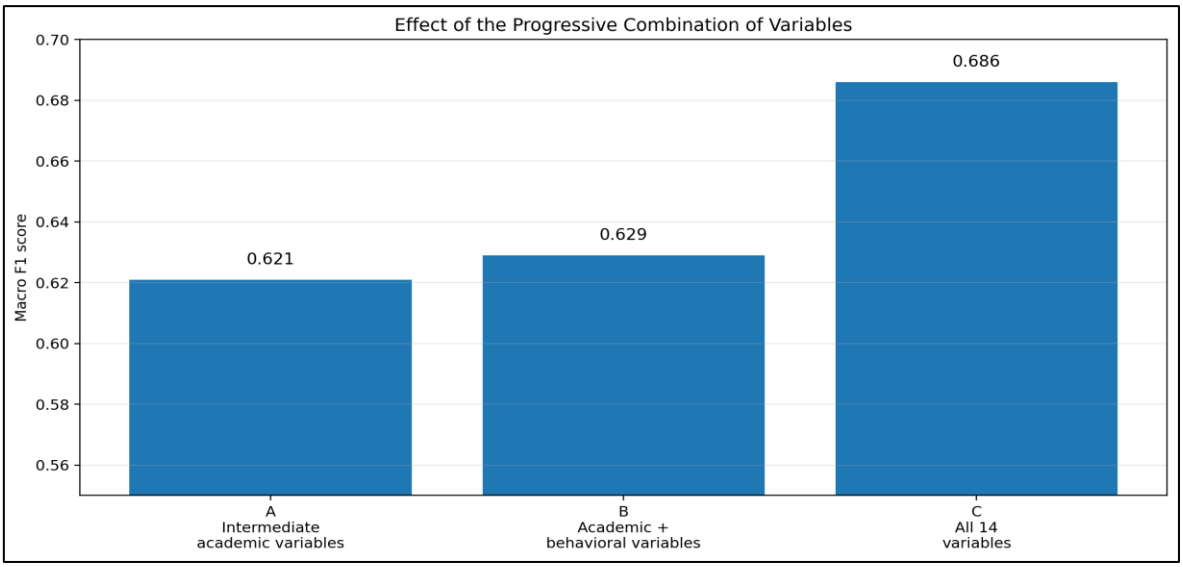

Fig 3 Effect of Progressive Combination of Variables on Macro F1-Score.
Source: Direct Graphical Output Generated by the Final Python/Jupyter Notebook.

➢ *Recommendations, Uncertainty Flags, and Classwise Diagnostics*

Each predicted class was linked to an intervention category. The mapping is intentionally graded: Risk and Fragile call for more immediate and structured support, Average calls for consolidation, and the higher classes guide the student toward maintenance or enrichment. Estimated probabilities did not impose a second decision; they only triggered educator review when the confidence or margin condition described above was met.

Table 6 Recommendation Framework and Human-Review Summary on Test Predictions

Profile	Proposed support	Cases	Mean conf.	Review
Risk	Rapid interview, diagnosis, tutoring, and remediation	77	0.773	13
Fragile	Targeted support, feedback, and mentoring	363	0.677	89
Average	Consolidation and periodic follow-up	628	0.671	177
Good	Maintenance and deepening	513	0.666	137
Very Good	Advanced activities	191	0.640	75
Excellence	Research project and peer tutoring	28	0.672	11

Source: Authors' Output from the Prediction-to-Recommendation Rule on the Independent Test Subset.

The final-model decision is not based on average metrics alone. The classwise diagnostic results show that the random forest performed most steadily for Fragile, Average, and Good profiles. Performance remains more cautious at the extremes, especially for Excellence, where the test subset contains only 37 observations. The AUC-ROC values show strong probability ranking across classes, but probability ranking should not be confused with perfect class assignment. Individual predictions must remain open to qualitative reading of the student record.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul246

Table 7 Class wise Performance and One-vs-Rest AUC-ROC of the Retained Random Forest

Class	Precision	Recall	F1-score	Support	AUC-ROC
Good	0.717	0.716	0.717	514	0.905
Excellence	0.679	0.514	0.585	37	0.985
Fragile	0.702	0.733	0.717	348	0.936
Average	0.713	0.740	0.727	605	0.888
Risk	0.792	0.649	0.713	94	0.978
Very Good	0.675	0.639	0.656	202	0.952
Macro average	0.713	0.665	0.686	1,800	0.941

Source: Authors' Class wise Report and One-vs-Rest AUC-ROC Output from the Independent Test Subset.

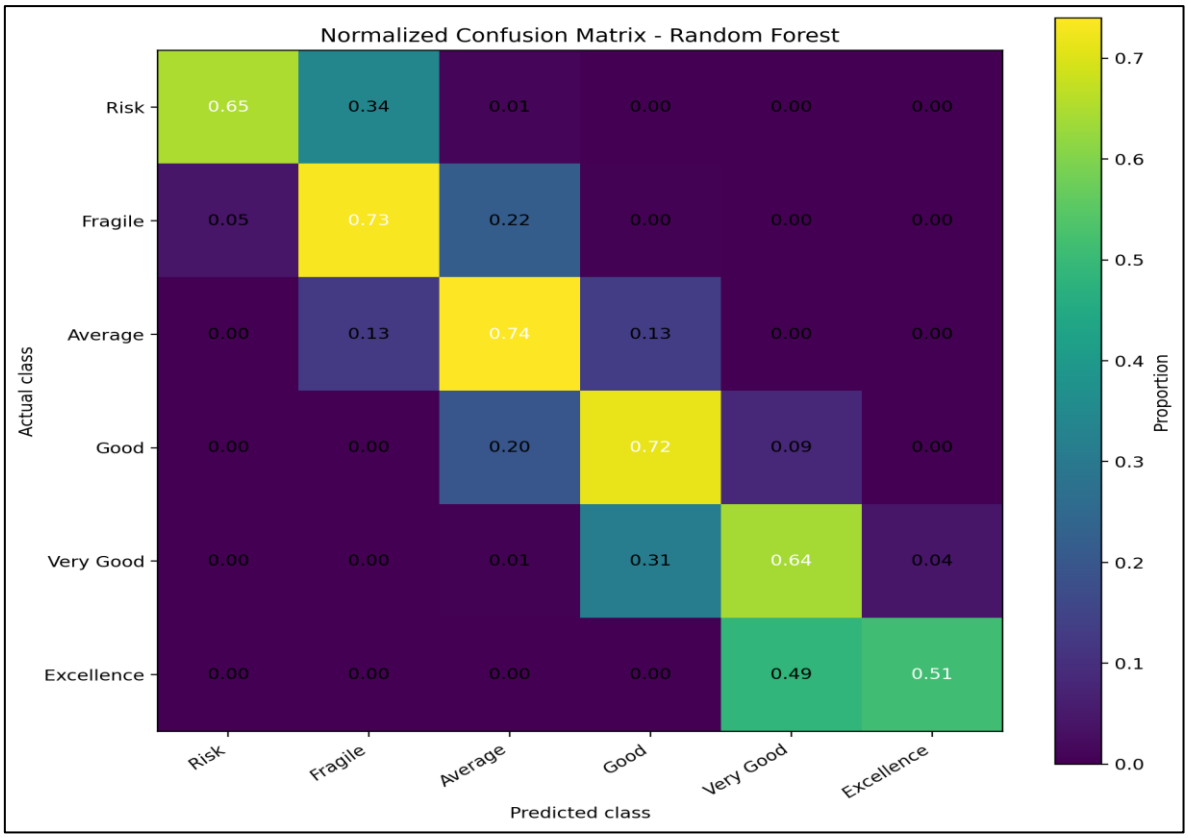

Fig 4 Normalized Confusion Matrix of the Final Random-Forest Model.
Source: Direct Graphical Output Generated by the Final Python/Jupyter Notebook.

The confusion matrix shows that most errors occur between neighbouring performance levels. Of the students actually in the Risk class, 64.9% were correctly identified and 34.0% were classified as Fragile. In practice, these categories should be treated as a shared zone of academic vigilance while maintaining different intervention intensity. Excellence had a recall of 0.514 and was mainly confused with Very Good, which is consistent with the small number of Excellence observations. The matrix therefore supports a cautious interpretation: the output is a signal for dialogue and follow-up, not a definitive label.

➢ *Discussion*

The study links early multiclass prediction with a practical academic recommendation framework. Whereas many studies reduce performance to a pass/fail outcome, the six-level structure used here represents a more gradual academic trajectory and connects it with differentiated pedagogical responses. This approach is consistent with the view that learning analytics should support learning rather than merely rank students [9], [10].

Several design choices strengthen the internal validity of the experiment: stratified separation of training and test data, cross-validation restricted to the training subset, pipeline-based preprocessing, and final evaluation on an independent sample. Together, these measures reduce the risk of information leakage and offer a more realistic estimate of out-of-sample performance [17]. The prespecified use of macro F1 is especially important because a high accuracy could hide weak recognition of Risk and Excellence, the minority categories in the corpus.

Random forest was retained because it combined the strongest mean macro F1 during five-fold cross-validation with a direct interpretation pathway through permutation importance. This is consistent with its ability to capture nonlinear interactions in heterogeneous predictors [14]. The

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul246

multilayer perceptron achieved higher overall accuracy and macro AUC-ROC, while logistic regression remained attractive where balanced class sensitivity and a simpler linear benchmark are prioritised. The selected model should therefore be understood as the most useful option for this experimental objective, not as a model that will dominate every alternative in every setting.

The most influential variables emphasise the value of interim assessment and continuity in the student trajectory. Continuous assessments, midterm average, and previously validated credits contained more predictive information than behavioural indicators taken alone. This does not establish a causal hierarchy; it shows only that these variables helped the model separate performance profiles more effectively in the present corpus. The scenario analysis reinforces the same point: early warnings become more informative when current-semester evidence, prior progress, contextual information, and regularity markers are considered together.

The results are consistent with studies that regard predictive analytics as a tool for early warning and targeted follow-up [7], [8]. They also require restraint in how categories are communicated. Risk is sometimes confused with Fragile and Excellence with Very Good, but these are predominantly neighbouring classes. Such a pattern supports the use of predictions as prompts for conversation and support rather than fixed identities assigned to students.

V. CONCLUSION

This study evaluated an early academic-performance prediction model and translated its outputs into support-oriented recommendations. Using 9,000 student-semester observations and fourteen predictors available by the middle of the semester, four supervised algorithms were compared with measures suited to an imbalanced multiclass problem. Random forest was retained after five-fold cross-validation and confirmation on the independent test set. It achieved macro F1 of 0.686, macro precision of 0.713, and macro One-vs-Rest AUC-ROC of 0.941.

Continuous assessments, midterm average, and previously validated credits emerged as the most informative signals. The results further show that combining academic, behavioural, historical, and contextual information is more useful than relying on an isolated grade. The study's main contribution is the connection between a six-level prediction and a graduated recommendation rule. The predicted class proposes a support option, while low-confidence or low-margin cases are explicitly referred to human review. The model does not replace teachers or academic authorities; it makes early signals more visible so that support can be considered before difficulties become entrenched.

Institutional use should be gradual. A pilot on anonymized cohorts can be followed by validation across academic years and programmes. Teachers, academic leaders, and information-system staff should agree on data elements, access rights, alert thresholds, corrective procedures, and intervention options. Any full deployment should connect the validated model to secure data processes and be evaluated for fairness, comprehensibility, and educational impact.

VI. LIMITATIONS AND FUTURE RESEARCH

The harmonized dataset was prepared for experimental purposes and does not substitute for a longitudinal institutional database integrated across programmes and academic years. Its generalisability should therefore be examined on real anonymized cohorts and across successive semesters. The small number of Excellence cases also reduces the stability of estimates for that class. In addition, the effectiveness of the proposed recommendations was not tested in a live intervention setting. Future work should assess not only predictive performance but also whether tutoring, remediation, feedback, and enrichment measures improve subsequent student progress.

Ethical and operational questions remain inseparable from this perspective. Student data should be minimised, protected, and used for a defined educational purpose; transparency, confidentiality, and human oversight are essential [11, 12, 13]. Future institutional implementation should therefore preserve traceability and provide a documented pathway for professional review at each stage.

REFERENCES

[1]. P. Long and G. Siemens, "Penetrating the fog: Analytics in learning and education," EDUCAUSE Review, vol. 46, no. 5, pp. 30–40, 2011.
[2]. G. Siemens and R. S. J. d. Baker, "Learning analytics and educational data mining: Towards communication and collaboration," in Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, 2012, pp. 252–254, doi: 10.1145/2330601.2330661.
[3]. C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," WIREs Data Mining and Knowledge Discovery, vol. 10, no. 3, e1355, 2020, doi: 10.1002/widm.1355.
[4]. R. S. J. d. Baker and P. S. Inventado, "Educational data mining and learning analytics," in Learning Analytics: From Research to Practice, J. A. Larusson and B. White, Eds. Springer, 2014, pp. 61–75, doi: 10.1007/978-1-4614-3305-7_4.
[5]. S. B. Kotsiantis, C. J. Pierrakeas, and P. E. Pintelas, "Predicting students' performance in distance learning using machine learning techniques," Applied Artificial Intelligence, vol. 18, no. 5, pp. 411–426, 2004, doi: 10.1080/08839510490442058.
[6]. P. Cortez and A. M. G. Silva, "Using data mining to predict secondary school student performance," in Proceedings of the 5th Annual Future Business Technology Conference, 2008, pp. 5–12.
[7]. S. M. Jayaprakash, E. W. Moody, E. J. M. Lauría, J. R. Regan, and J. D. Baron, "Early alert of academically at-risk students: An open source analytics initiative," Journal of Learning Analytics, vol. 1, no. 1, pp. 6–47, 2014.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul246

[8]. D. Ifenthaler and J. Y.-K. Yau, "Using learning analytics to support study success in higher education: A systematic review," Educational Technology Research and Development, vol. 68, pp. 1961–1990, 2020, doi: 10.1007/s11423-020-09788-z.
[9]. D. Gašević, S. Dawson, and G. Siemens, "Let's not forget: Learning analytics are about learning," TechTrends, vol. 59, no. 1, pp. 64–71, 2015, doi: 10.1007/s11528-014-0822-x.
[10]. G. Siemens, "Learning analytics: The emergence of a discipline," American Behavioral Scientist, vol. 57, no. 10, pp. 1380–1400, 2013, doi: 10.1177/0002764213498851.
[11]. S. Slade and P. Prinsloo, "Learning analytics: Ethical issues and dilemmas," American Behavioral Scientist, vol. 57, no. 10, pp. 1510–1529, 2013, doi: 10.1177/0002764213479366.
[12]. N. Sclater, "Developing a code of practice for learning analytics," Journal of Learning Analytics, vol. 3, no. 1, pp. 16–42, 2016, doi: 10.18608/jla.2016.31.3.
[13]. UNESCO, Recommendation on the Ethics of Artificial Intelligence. Paris, France: UNESCO, 2021.
[14]. S. Kaufman, S. Rosset, C. Perlich, and O. Stitelman, "Leakage in data mining: Formulation, detection, and avoidance," ACM Transactions on Knowledge Discovery from Data, vol. 6, no. 4, Art. no. 15, 2012, doi: 10.1145/2382577.2382579.
[15]. M. Kuhn and K. Johnson, Feature Engineering and Selection: A Practical Approach for Predictive Models. Boca Raton, FL: CRC Press, 2019.
[16]. L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[17]. T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. New York, NY: Springer, 2009.
[18]. T. Fawcett, "An introduction to ROC analysis," Pattern Recognition Letters, vol. 27, no. 8, pp. 861–874, 2006, doi: 10.1016/j.patrec.2005.10.010.
[19]. T. Kluyver et al., "Jupyter Notebooks – a publishing format for reproducible computational workflows," in Positioning and Power in Academic Publishing: Players, Agents and Agendas, F. Loizides and B. Schmidt, Eds. IOS Press, 2016, pp. 87–90, doi: 10.3233/978-1-61499-649-1-87.
[20]. W. McKinney, "Data structures for statistical computing in Python," in Proceedings of the 9th Python in Science Conference, 2010, pp. 56–61, doi: 10.25080/Majora-92bf1922-00a.
[21]. C. R. Harris, K. J. Millman, S. J. van der Walt, et al., "Array programming with NumPy," Nature, vol. 585, pp. 357–362, 2020, doi: 10.1038/s41586-020-2649-2.
[22]. F. Pedregosa, G. Varoquaux, A. Gramfort, et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
[23]. J. D. Hunter, "Matplotlib: A 2D graphics environment," Computing in Science & Engineering, vol. 9, no. 3, pp. 90–95, 2007, doi: 10.1109/MCSE.2007.55.